CRENoS

Centro Ricerche Economiche Nord Sud
Università degli Studi di Cagliari
Università degli Studi di Sassari

THE PROPERTIES OF SOME GOODNESS OF FIT TESTS

Gianna Boero
Jeremy Smith
Kenneth F. Wallis

CONTRIBUTI DI RICERCA
02/09

Gianna Boero
(University of Cagliari, Crenos, and University of Warwick)
boero@unica.it

Jeremy Smith
(University of Warwick)
Jeremy.Smith@warwick.ac.uk

Kenneth F. Wallis
(University of Warwick)
K.F.Wallis@warwick.ac.uk

THE PROPERTIES OF SOME GOODNESS-OF-FIT TESTS

Abstract
The properties of Pearson's goodness-of-fit test, as used in density forecast evaluation, income distribution analysis and elsewhere, are analysed. The components-of-chi-squared or "Pearson analog" tests of Anderson (1994) are shown to be less generally applicable than was originally claimed. For the case of equiprobable classes, where the general components tests remain valid, a Monte Carlo study shows that tests directed towards skewness and kurtosis may have low power, due to differences between the class boundaries and the intersection points of the distributions being compared. The power of individual component tests can be increased by the use of nonequiprobable classes.

JEL classification: C12, C14

Keywords: Pearson's Goodness-of-fit test; Component tests; Distributional assumptions; Monte Carlo; Normality; Nonequiprobable partitions.

Acknowledgement: We would like to thank Gordon Anderson for extremely helpful comments.

October 2002

1. Introduction

Many areas of economics require the comparison of distributions. Analysis of the distribution of income has a history that extends over more than a century, while a new area of application is the evaluation of density forecasts, that is, forecasts expressed as estimates of the complete probability distribution of possible future outcomes. The statistical problem in all such applications is to assess the degree of correspondence or goodness of fit between observed data and a hypothesised distribution. The two classical nonparametric approaches to testing goodness of fit are based on grouping data into classes or calculating the sample distribution function and in each case comparing observation to hypothesis. Pearson's chi-squared test and the Kolmogorov-Smirnov test are the best-known procedures in the two respective cases, surveyed by Stuart, Ord and Arnold (1999, Ch. 25). Both have recent application to the evaluation of density forecasts of inflation, by Diebold, Tay and Wallis (1999) and Wallis (2002).

Anderson (1994) presents a rearrangement of the chi-squared goodness-of-fit statistic to provide more information on the nature of departures from the hypothesised distribution, in respect of specific features of the empirical distribution such as its location, scale and skewness. An application of this components-of-chi-squared or "Pearson analog" test to the comparison of income distributions is given by Anderson (1996). It is also used in density forecast evaluation by Wallis (2002), noted above, and by Boero and Marrocu (2002), who compare density forecasts of exchange rates from a range of competing models.

A formal derivation of the components test is presented in this paper, and it is shown that some of Anderson's claims for the generality of the test are not correct. In the more restricted case in which the test remains valid, we proceed to a Monte Carlo study of its properties. The experimental design is motivated by applications of density forecasting in macroeconomics and finance. We simulate data from various distributions that are either used directly in real-time forecasting or that capture well-known features of many financial time series, namely their skewness and excess kurtosis. The Jarque and Bera (1980) test is also included for comparative purposes.

The paper proceeds as follows. Section 2 discusses the chi-squared goodness-of-fit test and analyses the properties of the decomposition proposed by Anderson (1994). Section 3 outlines the distributions used to generate artificial data that exhibits either skewness or kurtosis. The results of the Pearson analog test to detect departures from normality using both equiprobable and nonequiprobable splits for the partition points are reported in section 4. Finally, in section 5 we summarise the main results and make some concluding remarks.

2. The chi-squared goodness-of-fit test and its components

Pearson's classical goodness-of-fit test proceeds by dividing the range of the variable into k mutually exclusive classes and comparing the probabilities of outcomes falling in these classes given by the hypothesised distribution with the observed relative frequencies. With class probabilities $p_i > 0,\ i = 1,\ldots,k,\ \Sigma p_i = 1$ and class frequencies $n_i > 0,\ i = 1,\ldots,k,\ \Sigma n_i = n$, the test statistic is

$$X^2 = \sum_{i=1}^{k} \frac{(n_i - np_i)^2}{np_i}.$$

This has a limiting c^2 distribution with $k-1$ degrees of freedom if the hypothesised distribution is correct.

The asymptotic distribution of the test statistic rests on the asymptotic k-variate normality of the multinomial distribution of the observed frequencies. Placing these in the $k \times 1$ vector $\boldsymbol{x}$, under the null hypothesis this has mean vector $m = (np_1, np_2, \ldots, np_k)'$ and covariance matrix

$$\boldsymbol{V} = n\begin{bmatrix} p_1(1-p_1) & -p_1p_2 & \dots & -p_1p_k \\ -p_2p_1 & p_2(1-p_2) & \dots & -p_2p_k \\ \vdots & \vdots & & \vdots \\ -p_kp_1 & -p_kp_2 & \dots & p_k(1-p_k) \end{bmatrix}$$

This matrix is singular, with rank $k-1$: note that each column (row) has sum zero. Defining the generalized inverse $\boldsymbol{V}^-$, the quadratic form $(\boldsymbol{x}-m)'\boldsymbol{V}^-(\boldsymbol{x}-m)$ then has the χ^2 distribution with k–1 degrees of freedom (Pringle and Rayner, 1971, p.78).

In his derivation and application of the components test Wallis (2002) assumes that the classes are equiprobable, which is often recommended to improve the power properties of the overall test. In this case $p_i = 1/k,\ i = 1,\dots,k$ and

$$\boldsymbol{V} = (n/k)\left[\boldsymbol{I} - \boldsymbol{ee}'/k\right],$$

where $\boldsymbol{e}$ is a k×1 vector of ones. Since the matrix in square brackets is symmetric and idempotent it coincides with its generalized inverse, and the chi-squared statistic is equivalently written

$$X^2 = \sum_{i=1}^{k} \frac{(n_i - n/k)^2}{(n/k)}$$
$$= (\boldsymbol{x}-m)'[\boldsymbol{I} - \boldsymbol{ee}'/k](\boldsymbol{x}-m)/(n/k)$$

(note that $\boldsymbol{e}'(\boldsymbol{x}\text{-}m) = 0)$. There exists a $(k-1)\times k$ transformation matrix $\boldsymbol{A}$, such that

$$\boldsymbol{AA}' = \boldsymbol{I},\ \ \boldsymbol{A}'\boldsymbol{A} = \left[\boldsymbol{I} - \boldsymbol{ee}'/k\right]$$

(Rao and Rao, 1998, p.252). Hence defining $\boldsymbol{y} = \boldsymbol{A}(\boldsymbol{x}-m)$ the statistic can be written as an alternative sum of squares

$$X^2 = \boldsymbol{y}'\boldsymbol{y}/(n/k)$$

where the $k-1$components $y_i^2/(n/k)$ are independently distributed as c^2 with one degree of freedom under the null hypothesis.

To construct the matrix $\boldsymbol{A}$ we consider Hadamard matrices, which are square matrices whose elements are 1 or -1 and whose columns are orthogonal: $\boldsymbol{H}'\boldsymbol{H} = k\boldsymbol{I}$. For k equal to a power of 2, we begin with the basic Hadamard matrix

$$\boldsymbol{H}_2 = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$$

and form Kronecker products

$$\boldsymbol{H}_4 = \boldsymbol{H}_2 \otimes \boldsymbol{H}_2,\ \boldsymbol{H}_8 = \boldsymbol{H}_4 \otimes \boldsymbol{H}_2 \text{ and } \boldsymbol{H}_{16} = \boldsymbol{H}_8 \otimes \boldsymbol{H}_2.$$

Deleting the first row of 1s and dividing by $\sqrt{k}$ then gives the required matrix $\boldsymbol{A}$, that is, $\boldsymbol{H}$ is partitioned as

$$\boldsymbol{H} = \begin{bmatrix} e' \\ \sqrt{k}A \end{bmatrix}.$$

With $k = 4$, and rearranging rows, we have

$$\boldsymbol{A} = \frac{1}{2}\begin{bmatrix} 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 \\ 1 & -1 & 1 & -1 \end{bmatrix}$$

and the three components of the test focus in turn on departures from the null distribution with respect to location, scale and skewness. Location shifts refer to the median, and scale shifts to the inter-quartile range, while the third component detects possible asymmetries, that is, shifts between the first and third quarters and the second and fourth quarters of the distribution.

Taking $k = 8$ allows a fourth component related to kurtosis to appear, although the remaining three components are difficult to relate to characteristics of the distribution:

$$\boldsymbol{A} = \frac{1}{\sqrt{8}}\begin{bmatrix} 1 & 1 & 1 & 1 & -1 & -1 & -1 & -1 \\ 1 & 1 & -1 & -1 & -1 & -1 & 1 & 1 \\ 1 & 1 & -1 & -1 & 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 & 1 & -1 & -1 & 1 \\ 1 & -1 & 1 & -1 & 1 & -1 & 1 & -1 \\ 1 & -1 & 1 & -1 & -1 & 1 & -1 & 1 \\ 1 & -1 & -1 & 1 & -1 & 1 & 1 & -1 \end{bmatrix}.$$

In this case the first four individual components of the test statistic are identified with particular features of the distribution, and the remainder is independently distributed as χ^2 with three degrees of freedom under the null hypothesis. Note that the first three components when $k = 8$ coincide with the (only) three components when $k = 4$.

Returning to the general case of unequal class probabilities we write the covariance matrix as

$$\boldsymbol{V} = n\left[\boldsymbol{P} - \boldsymbol{pp}'\right]$$

where $\boldsymbol{P} = diag(p_1, p_2, \ldots, p_k)$ and $\boldsymbol{p} = (p_1, p_2, \ldots, p_k)'$. Tanabe and Sagae (1992) give the result that

$$\left[\boldsymbol{P} - \boldsymbol{pp}'\right]^{-} = (\boldsymbol{I} - \boldsymbol{ee}'/k)\boldsymbol{P}^{-1}(\boldsymbol{I} - \boldsymbol{ee}'/k)$$

hence the test statistic can again be written in terms of the transformed variables $\boldsymbol{y} = \boldsymbol{A}(\boldsymbol{x} - m)$, as

$$X^2 = \boldsymbol{y}'\boldsymbol{A}\boldsymbol{P}^{-1}\boldsymbol{A}'\boldsymbol{y}/n.$$

The covariance matrix of $\boldsymbol{y}$ is $E(\boldsymbol{yy}') = \boldsymbol{AVA}'$, whose inverse $\boldsymbol{AP}^{-1}\boldsymbol{A}'/n$ appears in the above quadratic form: this can be checked by multiplying out, noting that $\boldsymbol{e}'\left[\boldsymbol{P} - \boldsymbol{pp}'\right] = \boldsymbol{0}$ and $\boldsymbol{AP}^{-1}\boldsymbol{p} = \boldsymbol{Ae} = \boldsymbol{0}$. The diagonal elements give the variance of y_i, $i = 1,\ldots,k-1$, as $s_i^2 = n\left(1 - k(\boldsymbol{a}_i'\boldsymbol{p})^2\right)/k$ where $\boldsymbol{a}_i'$ is the ith row of $\boldsymbol{A}$, corresponding to expressions given by Anderson (1994, p.267) for the first four elements when $k = 8$. However, the matrix is not diagonal – the y_i's are correlated – hence this quadratic form does not reduce to a simple sum of squares in the case of unequal p_i. Equivalently, terms of the form y_i^2 / s_i^2 are not independently distributed as c^2 in this general case, contrary to Anderson's claim. It remains the case, however, that the marginal distribution of an individual y_i^2 / s_i^2 is c^2 with one degree of freedom.

Anderson (1994) goes on to argue that the power of the component tests to detect departures of the alternative distribution from the null distribution depends on the closeness of the intersection points of the two distributions to the location of the class boundaries. In the equiprobable case the class boundaries are the appropriate quantiles, and moving them to improve test performance clearly alters the class probabilities. In practical examples this is often done in such a way that the resulting class probabilities are symmetric, that is, $p_1 = p_k,\ p_2 = p_{k-1}, \ldots,$ which is an interesting special case. Now the odd-numbered and even-numbered components of the chi-squared statistic are orthogonal. Hence focussing only on location and scale, for example, a joint test can be based on $y_1^2 / s_1^2 + y_2^2 / s_2^2$, distributed as c_2^2 under the null; similarly, focussing only on skewness and kurtosis, a joint test can be based on $y_3^2 / s_3^2 + y_4^2 / s_4^2$. As in the more general case, however, these two test statistics are not independent, equivalently their sum is not distributed as c^2 with four degrees of freedom.

This observation helps to explain some of the simulation results of Anderson (2001), who includes components-of-chi-squared tests in a comparison of a range of tests for location and scale problems. With four non-equiprobable classes whose boundaries are placed symmetrically around the mean, the class probabilities are symmetric in the sense of the previous paragraph if the null distribution is symmetric, but not otherwise. The components considered are the individual location and scale components, which in general are c_1^2; a "joint" test based on their sum, as in the above example, which is c_2^2 in the symmetric case but not in general; and a "general" test defined as the sum of the three components, which is no longer c_3^2 even in the symmetric but non-equiprobable case. Anderson (2001, p.25) reports that the power and consistency properties of the various components tests are good, "with the exception of the general and joint tests under the asymmetric distribution." Our analysis shows that in these circumstances the null distribution is not c^2, contrary to what is assumed in the simulation study, hence the problems noted are the result of comparing test statistics to inappropriate critical values.

3. The Monte Carlo experiments

Our Monte Carlo experiments consider the power of the overall goodness-of-fit test and its component tests to detect departures from a standard normal distribution, in the presence of either skewness or kurtosis. We consider three skewed distributions and three kurtotic distributions, as follows.

3.1 Skewness

(i) The Ramberg distribution (see Ramberg *et al*., 1979), is a flexible form expressed in terms of its cumulative probabilities. The Ramberg quantile and density functions have the form:

$$R(p) = l_1 + \left[p^{l_3} - (1-p)^{l_4} \right] / \, l_2$$

$$f(x) = f[R(p)] = l_2 \left[l_3 p^{l_3 - 1} + l_4 (1-p)^{l_4 - 1} \right]$$

with $0 < p < 1$ being the cumulative probability, $R(p)$ the corresponding quantile, and $f[R(p)]$ the density corresponding to $R(p)$. Of the four parameters, l_1 is the location parameter, l_2 the scale parameter, and l_3 and l_4 are shape parameters. For the present purpose we choose their values such that $E(X) = 0, V(X) = 1$, Skewness={0.00, 0.05, 0.10, …, 0.90} and

Kurtosis =3. The median is then in general non-zero; it is an increasing function of the skewness. The non-zero median gives power to the goodness-of-fit test due to the contribution of the first component test. In order to concentrate on the effect of skewness alone we shift the distribution by the empirically calculated median.

(ii) The two-piece normal distribution (see Wallis, 1999), is used by the Bank of England and the Sveriges Riksbank in presenting their density forecasts of inflation. The probability density function is

$$f(x) = \begin{cases} \left[\sqrt{2p}\,(s_1 + s_2)/2)\right]^{-1} \exp\left[-\left(x-m\right)^2 / 2s_1^2\right] & x \le m \\ \\ \left[\sqrt{2p}\,(s_1 + s_2)/2)\right]^{-1} \exp\left[-\left(x-m\right)^2 / 2s_2^2\right] & x \ge m \end{cases}$$

.

The distribution is positively skewed if $s_2^2 > s_1^2$, and is leptokurtic if $s_1 \neq s_2$. As in the Ramberg distribution the median is an increasing function of skewness and we again shift the distribution, to ensure a theoretical median of zero. In our simulations we consider combinations of (s_1, s_2) that yield $V(X) = 1$ and the same range of skewness coefficients, namely {0.00, 0.05, 0.10, …, 0.90}.

(iii) Our third distribution is the data generating process used by Anderson (1994), namely

$$x = \begin{cases} (z/(1+d)) & z < 0 \\ z(1+d) & \text{otherwise} \end{cases}$$

where $z \sim N(0,1)$. Since skewness≈2×*d* we set *d*={0.00, 0.025, …, 0.45}. The mean, variance and kurtosis of this distribution are all increasing functions of *d*, although the median is zero. The transformation is discontinuous at zero, hence the probability density function has a central singularity, unlike the two-piece normal distribution.

3.2 Kurtosis

(i) In the Ramberg distribution we choose values of the four parameters such that $E(X)=0, V(X)=1$, Skewness=0 and Kurtosis={2.0, 2.4, 2.8, 3.0, 3.2. 3.6, 4.0, 4.4, 4.8, 5.2, 5.6, 6.0, 6.4, 6.8, 7.2}.

(ii) The *t*-distribution is widely used to represent the excess kurtosis of many financial time series. We scale it to have unit variance, and choose degrees of freedom n={5, 6, 7, 8, 9, 10, 16, 24, 34, ∞}, where kurtosis is given as $3(u-2)/(u-4)$ for $u>4$

(iii) The second data generating process used by Anderson (1994) is

$$x = z\left(|z|^{q}\right)(1+t)$$

where $z \sim N(0,1)$ and t is a variance-shifting nuisance parameter. We take all combinations of q and t that give $V(X)=1$ and kurtosis is approximately equal to those values used in the Ramberg distribution.

4. The Monte Carlo results

The results of the Monte Carlo experiments reported in this paper are based on 1000 replications for sample sizes, n=25, 50, 75, 100, 150, 250, 350. Nonequiprobable classes are considered from time to time in the following discussion, and our convention is to present class boundaries implicitly, as the appropriate percentage points of the relevant cumulative distribution function, F, the boundaries being the corresponding x-coordinates. We denote the value of the cdf at the upper boundary of the jth class as F_{j}. The first and last classes are open-ended, thus with F_0 =0 and F_k =1 the class probabilities satisfy, $p_j = F_j - F_{j-1}$, j=1,…,k, and a class configuration is reported as the set $\{F_1,\ldots,F_{k-1}\}$. For equiprobable classes, $F_j=j/k$.

While for the overall goodness-of-fit test the choice of k is important, the individual component tests do not depend on k, once k is large enough to define them. Thus when k=4 three components are defined, and these are unchanged when k=8, assuming that the eight classes are obtained by dividing each of the original four classes into two, without moving the class boundaries. That is, F_2, F_4 and F_6 when k=8 are identical to F_1, F_2, F_3 when k=4. This is obviously the case when classes are equally probable, and it is a reasonable presumption otherwise. Likewise the fourth component, that is defined when k=8, is unchanged when k=16, so here we use k=8 so that all components up to and including that related to kurtosis can be calculated.

In all cases we compare the results of the Pearson chi-squared component tests with the corresponding component, either skewness or kurtosis, of the test for normality of Jarque and Bera (1980). These are based on the ratios of the sample third and fourth moments to their standard errors, the latter being calculated via the higher moment relationships of the normal distribution.

4.1 Skewness

Figure 1 plots the power of the Pearson component skewness (PCSk) test, based on equal partitions, when data are generated from the two-piece normal distribution. The power of the PCSk test increases with both the sample size and the degree of skewness in the underlying distribution. However, for all values of skewness and sample sizes the power of the PCSk test is dominated by the power of the skewness element of the Jarque-Bera test, denoted JBS. In particular, for n=100, the power of the PCSk test is 12.1 (16.8) for skewness=0.5 (0.7) compared with 48.5 (78.6) for the JBS test.

Our results regarding the Jarque-Bera test confirm with those reported in Anderson (1994), in that the JBS test is sensitive to non-normal kurtosis, and the kurtosis element of the Jarque-

Bera test, JBK, is sensitive to non-normal skewness. By contrast, all of the four component tests from the modified Pearson test are correctly sized (the exception being the PCK test which detects kurtosis for a skewed Ramberg distribution with kurtosis=3.0).

For the two-piece normal distribution, increasing skewness does induce increased kurtosis and the Pearson component kurtosis (PCK) test finds evidence of kurtosis as the skewness parameter increases. The power of the X^2 (denoted as X2 in the figures) test is markedly in excess of that of the PCSk test, due to the power coming from the residual component test. The large power contribution from the residual component test is clearly difficult to interpret. Figure 2 plots the power of the X^2, PCSk, PCK, the residual $\chi^2(3)$ (PCR) test as well as the JBS test, as a function of the skewness parameter for n=150.

Part of the possible explanation for the low power (see Anderson, 1994) of the PCSk test could be due to an inappropriate choice for the location of the partition points. Figure 3 plots the two-piece normal (skewness=0.5) and the N(0,1) distributions. We also plot the three important partition points for the PCSk test, assuming an equiprobable split at F_2=0.25, F_4=0.5 and F_6=0.75. The actual intersection points of the two distributions, which are much more in the tails of the N(0,1) distribution, do not coincide particularly well with these equiprobable splits.

Figure 4 plots the size of the skewness component test when we allow for nonequiprobable (but symmetrical) splits as: ($F_2/2$, F_2, $(0.5+F_2)/2$, 0.5, $(1-(0.5+F_2)/2)$, $(1- F_2)$, $(1- F_2/2)$), for F_2 taking values (0.15, 0.175, 0.2, …, 0.3). The figure (which also plots the 99% confidence intervals for a 5% nominal size test) shows that size is unaffected by the use of unequal partitions (with the exception of n=50, which is slightly over-sized). Figure 5 plots the power of the PCSk test as we vary F_2 for various values of skewness and n=150, although qualitatively similar pictures exist for all sample sizes. Power increases as F_2 becomes smaller. At F_2=0.15 (n=150 and skewness=0.5) power is 31.1% compared to 16.1% when using an equiprobable split, F_2=0.25. In general, the use of F_2=0.15 significantly improves the power of the skewness component of the test, with power nearly doubling for sample sizes from n=75 onwards. Despite this marked increase in power for the PCSk test, the JBS test still dominates (compare Figures 2 and 5).

While it is generally perceived that an equiprobable split maximises the power of the overall goodness-of-fit test, we find that the power of the X^2 test increases as F_2 falls and attains its maximum at F_2=0.15, irrespective of the sample size. For example, for the two-piece normal, with F_2=0.15 (n=150 and skewness=0.5) the power for the X^2 test is 29.2% compared to 22.1% for F_2=0.25. However, this power still compares unfavourably with that for the JB $\chi^2(2)$ test, which is 53.1%.

Results for the Ramberg distribution with skewness ranging from 0.0 to 0.85 are qualitatively very similar to those obtained with the two-piece normal distribution and are therefore not reported in detail here. Figure 6 plots the power of the X^2 and PCSk using equiprobable and nonequiprobable splits with F_2=0.15 (optimal value) as well as the JBS test. It is clear that the use of nonequiprobable splits again markedly increases the power of the PCSk and X^2 tests to detect skewness. Moreover we note that in the nonequiprobable case, while the JBS test unambiguously dominates the PCSk test for large sample sizes, with smaller sample sizes (n=25, 50, 75 and 100) the PCSk test actually outperforms the JBS test.

Figure 7 plots the power of the PCSk test when the data are generated from Anderson's skewed distribution, based on equiprobable splits. The power of the PCSk test to detect departures from normality for this distribution is greater for all values of skewness and at all sample sizes, relative to that observed for either the two-piece normal or Ramberg distributions (compare Figures 1, 6 and 7). For instance for the smallest sample size, n=25, power is in excess of 30% for skewness=0.85 using Anderson's distribution, but is only 8.4% for the two-piece normal and 9.8% for the Ramberg distribution. Moreover we find that the power of the PCSk test dominates that of the JBS test even with the use of equiprobable partition points.

Part of the explanation for the improved performance of the PCSk test with equiprobable splits is the insensitivity of the PCSk test to the location of the partition points for this distribution. The use of nonequiprobable splits increases the power of the PCSk test, but the increase is not as dramatic as with either the two-piece normal or Ramberg distributions – see Figure 8. The optimal value for F_2 is approximately 0.2 (much closer to the equiprobable F_2=0.25). For n=150 and skewness=0.5 the power increases from 66.7% (for F_2=0.25) to 74.9% (for F_2=0.2). The power of the X^2 test increases slightly as F_2 decreases, for example for skewness=0.5 and n=150 power equals 62.7% at F_2=0.2, compared with 59.4% at F_2=0.25.

4.2 Kurtosis

In this section we report the results of both the Ramberg distribution and Anderson's kurtotic distribution. The results for the scaled t-distribution are qualitatively very similar to those obtained with the Ramberg distribution and are not reported in detail.

In Figure 9 we plot the power of the PCK test when the data are generated from a Ramberg distribution with kurtosis. The use of equiprobable partitions renders the power of the PCK test very poor, even for very high degrees of kurtosis (=7.2). Looking at all the individual component tests the X^2 test does pick up non-normality, although this is almost entirely due to the scale component (PCS) test. These results are reported in Figure 10, for n=150, and are compared with those for the JBK test. The JBK dominates the PCK test, for all sample sizes and all values of kurtosis.

Despite the symmetry of the alternative hypothesis, the JBS statistic suggests substantial rejection of the null hypothesis. To reconcile these results, we recall that the relevant null hypothesis of the JB test is normality, not symmetry. Correct rejections of the null hypothesis are due to the inappropriateness of the normal distribution based standard error of the third moment. This is most easily seen for a t-distribution with degrees of freedom equal to six, where the sixth and higher moments do not exist.

The power of the PCK test is dependent on the closeness of the four important partition points (F_1, F_3, F_5, F_7) to the intersection points of the Ramberg distribution (with excess kurtosis) and a N(0,1). Figure 11 plots the Ramberg distribution (kurtosis=6) and a N(0,1), as well as the partition lines based on an equiprobable split. The four points of intersection of the two distributions are not close to the equiprobable partition points. However, the power of the PCS test is dependent on the closeness of the equiprobable partition points at F_2=0.25 and F_6=0.75, which actually correspond well with two of the intersection points observed. This suggests that a different choice of nonequiprobable (although symmetric) partition points could deliver more power for the PCK test.

Figure 12 plots the power of the PCK test for the Ramberg distribution (kurtosis=6 and n=150) as a function of the partition points, F_1 and F_3, where the partition points are (F_1, (F_1+F_3)/2, F_3, 0.5, (1-F_3), 1-(F_1+F_3)/2, (1-F_1)). The power function is very steep, with maximum power achieved at F_1=0.025 and F_3=0.275. For n=150, the power for the PCK test falls from 56.2% (at F_1=0.025 and F_3=0.275) to 7.1% (at F_1=0.125 and F_3=0.375). The shape of the power function is qualitatively similar for all values of T and kurtosis used in this analysis.

In Figure 13 we report the power of the X^2 test for F_1=0.025 and F_3=0.275. As we can see by comparing Figure 13 with Figure 10, the use of nonequiprobable splits affects the power of the X^2 test only slightly. For n=150 the power increases from 25.8% (equiprobable splits) to 30.0% (nonequiprobable splits). This result is due to the substantial fall observed in the power of the PCS test, which offsets the improvement in the PCK test. Despite the increase in power of the PCK test from the use of nonequiprobable splits, the JBK test continues to dominate the PCK test for all sample sizes, but only for kurtosis greater than 3.4.

Figure 14 plots the power of the PCK test when data are generated from Anderson's kurtotic distribution. In this case, the PCK test picks up kurtosis extremely well. These results suggest that the partition points and the intersection points (of the theoretical and empirical distributions) are much closer for Anderson's kurtotic distribution. Moreover, differently from the case of the Ramberg distribution all of the component tests, with the exception of the PCK test, have power approximately equal to the nominal 5% size. In this case the power of the PCK test dominates that of the JBS test.

Figure 15 plots the power function of the kurtosis component test as a function of F_1 and F_3 (for kurtosis=6 and n=100). The power of the kurtosis component test does vary over F_1 and F_3, with power at the optimal point F_1=0.025 and F_3=0.45. Although the slope of this function is flatter than that observed for the Ramberg distribution, between the optimal point and the equiprobable point (F_1=0.125 and F_3=0.375) the power falls from 99.5% to 67.4%. The power of the X^2 test at F_1=0.025 and F_3=0.45 is greater than that attained using an equiprobable split, for example, for n=100 and kurtosis=6.0 the power of the X^2 test increases from 67.4% to 99.5%.

5. Conclusions

In this paper we have derived the component tests from Pearson's goodness-of-fit test and have shown that in the general case of nonequiprobable splits the overall goodness-of-fit test cannot be derived as the sum of the component tests due to a non-diagonal covariance matrix for the component tests.

The Monte Carlo experiments have been designed to examine the power of the skewness and kurtosis component tests to detect departures from a standard normal distribution, in the presence of either skewness or kurtosis. Our results have revealed that the power of the component tests crucially depends on the location of the partition points. In particular, for a range of skewed and kurtotic distributions, a choice for the location of the partition points, away from the usual equiprobable split, could significantly improve the power of the component tests. The results have also shown that a strategy of maximising the component tests does not always maximise the power of the overall goodness-of-fit test. The use of nonequiprobable splits makes the power of the modified Pearson component test, at time, comparable to that attained from the appropriate Jarque-Bera component test.

The overall chi-squared goodness-of-fit statistic with k-1 degrees of freedom can be transformed into k-1 components which have the potential to offer more information about the nature of departures from the null hypothesis, provided that the k classes are equiprobable, which is recommended practice having the power of the overall test in mind. However, the power of the individual component tests may be low, but it can be improved by designing nonequiprobable classes more appropriate to the specific feature of interest. Now the decomposition into k-1 independent components is lost, and attention can only focus on a single characteristic of the distribution. Users of these tests should therefore take note of this trade-off. The choice is general or specific, as this approach cannot simultaneously provide a good test for both.

References

Anderson, G. (1994). Simple tests of distributional form. *Journal of Econometrics*, 62, 265-276.

Anderson, G. (1996). Nonparametric tests of stochastic dominance in income distributions. *Econometrica*, 64, 1183-1193.

Anderson, G. (2001). The power and size of nonparametric tests for common distributional characteristics. *Econometric Reviews*, 20, 1-30.

Bera, A. K. and Jarque, C. M. (1981), "Efficient tests for normality, homoskedasticity and serial independence of regression residuals", *Economics Letters*, 6, 255-259.

Boero, G. and Marrocu, E. (2002). The performance of non-linear exchange rate models: a forecasting comparison. *Journal of Forecasting*, forthcoming.

Diebold, F.X., Tay, A.S. and Wallis, K.F. (1999). Evaluating density forecasts of inflation: the Survey of Professional Forecasters. In *Cointegration, Causality, and Forecasting: A Festschrift in Honour of Clive W. J. Granger* (R.F. Engle and H. White, eds), pp.76-90. Oxford: Oxford University Press.

Noceti, P, Smith, J. and Hodges, S. (2000). An evaluation of tests of distributional forecasts. Discussion Paper No.102, Financial Options Research Centre, University of Warwick.

Pringle, R.M. and Rayner, A.A. (1971). *Generalized Inverse Matrices with Applications to Statistics*. London: Charles Griffin.

Ramberg, J.S., Dudewicz, E.J., Tadikamalla, P.R. and Mykytka, E. (1979). A probability distribution and its uses in fitting data. *Technometrics*, 21, 201-214.

Rao, C.R. and Rao, M.B. (1998). *Matrix Algebra and its Applications to Statistics and Econometrics.* Singapore: World Scientific Publishing Co.

Stuart, A., Ord, J.K. and Arnold, S. (1999). *Kendall's Advanced Theory of Statistics*, 6th ed., vol. 2A. London: Edward Arnold.

Tanabe, K. and Sagae, M. (1992). An exact Cholesky decomposition and the generalized inverse of the variance-covariance matrix of the multinomial distribution, with applications. *Journal of the Royal Statistical Society* B, 54, 211-219.

Wallis, K.F. (1999). Asymmetric density forecasts of inflation and the Bank of England's fan chart. *National Institute Economic Review*, No. 167, 106-112.

Wallis, K.F. (2002). Chi-squared tests of interval and density forecasts, and the Bank of England's fan charts. *International Journal of Forecasting*, forthcoming

Figure 1: Power of PCSk test as a function of skewness: Two-piece normal

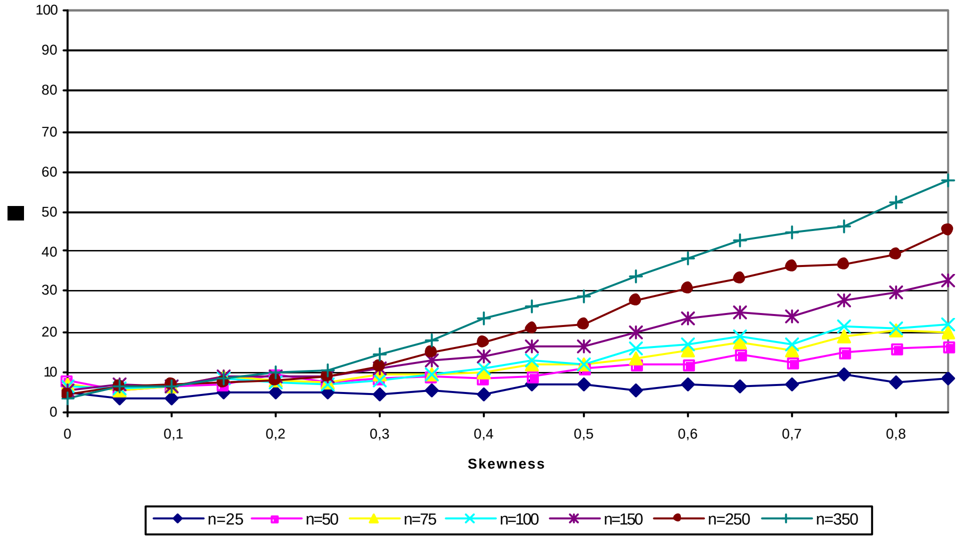

Figure 2: Power of the X^2, PCSk, PCR and JBS test for equiprobable splits: Two-piece normal (n=150)

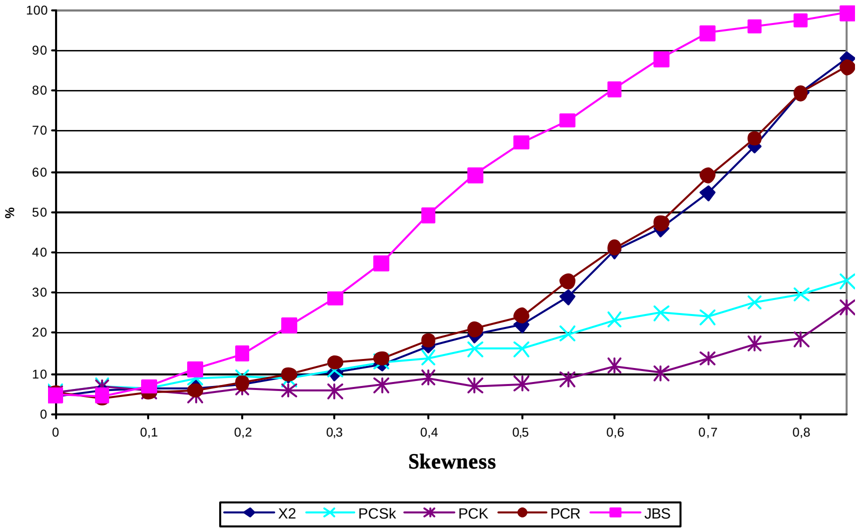

Figure 3: Intersection and partition points for PCSk test: Two-piece normal (skewness=0.5) and N(0,1)

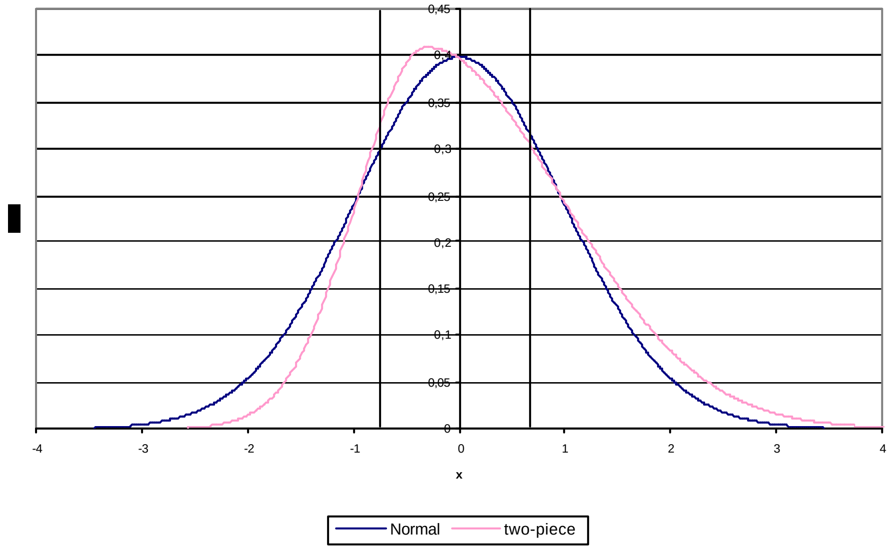

Figure 4: Size of PCSk test for non-equiprobable splits

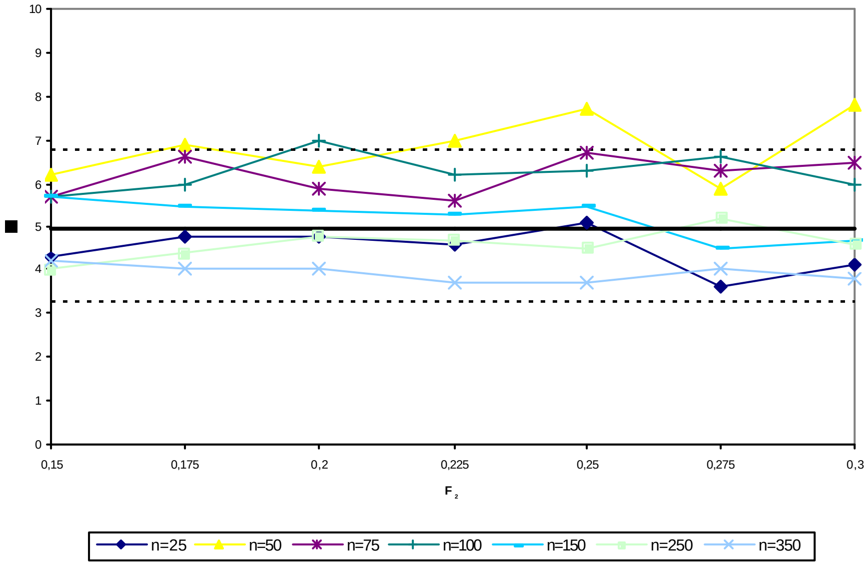

Figure 5: Power of the PCSk test for non-equiprobable splits: Two-piece normal (n=150)

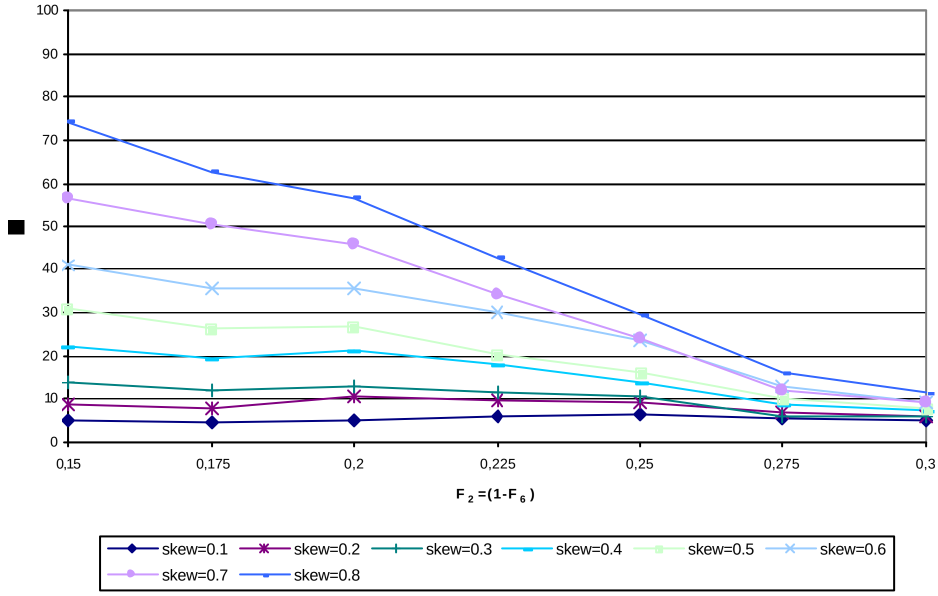

Figure 6: Power of the X^2, PCSk, and JBS tests: Ramberg distribution (n=150)

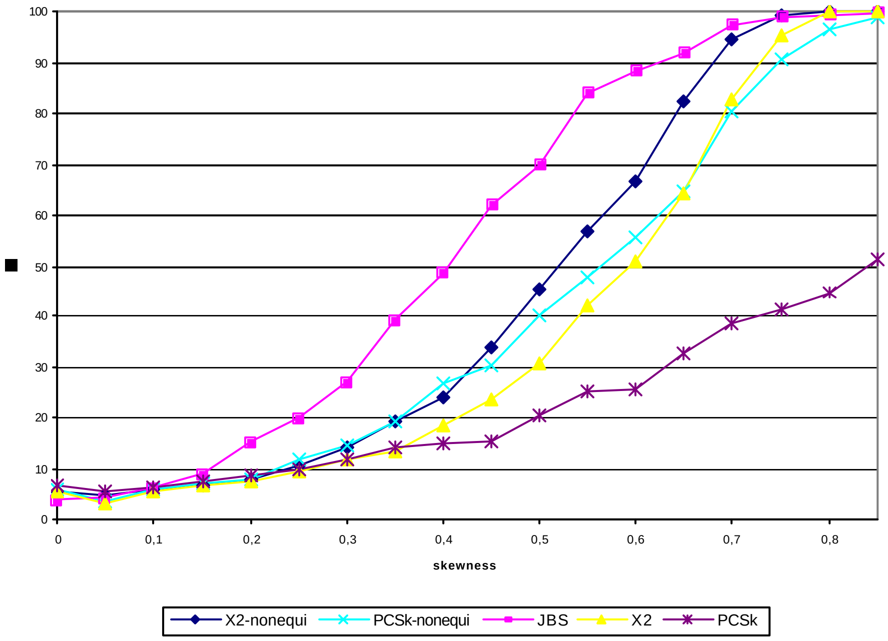

Figure 7: Power of PCSk test as a function of skewness for equiprobable splits: Anderson's skewed distribution

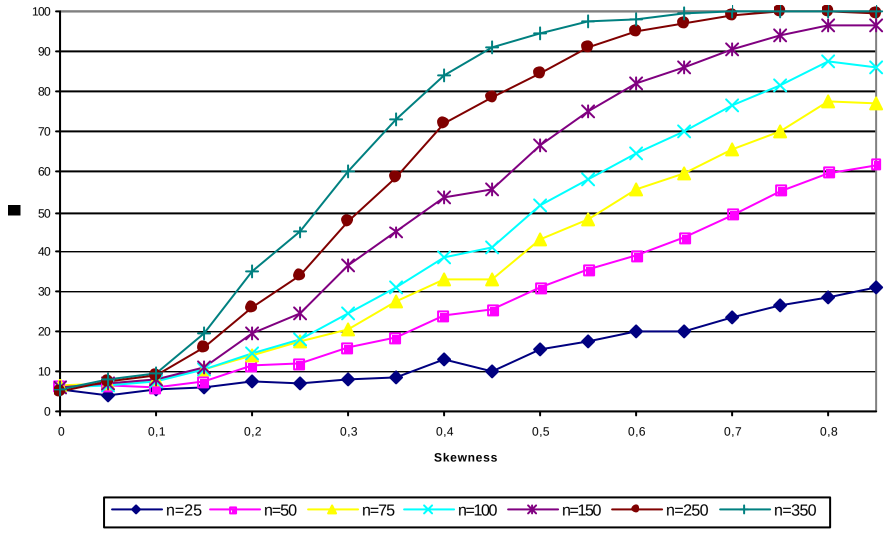

Figure 8: Power of PCSk test for non-equiprobable splits: Anderson's skewed distribution (n=150)

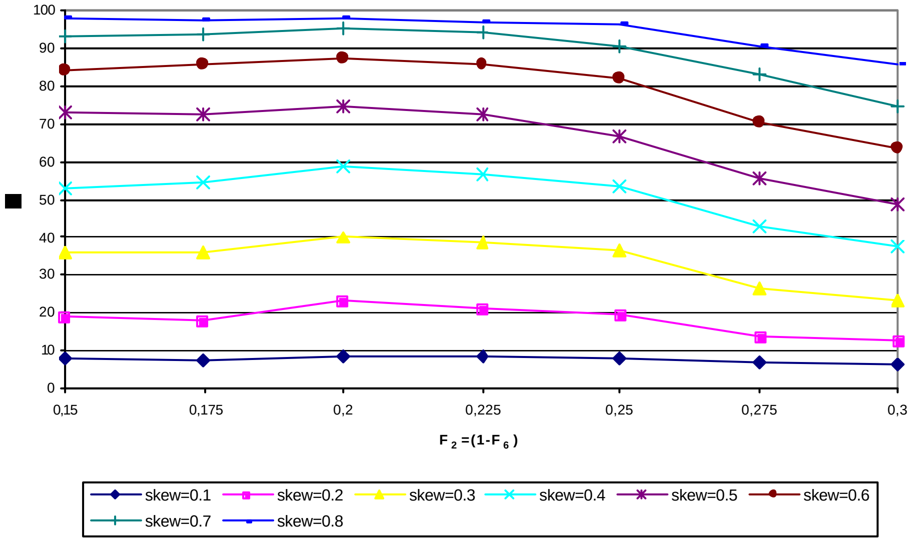

Figure 9: Power of PCK test as a function of kurtosis for equiprobable splits: Ramberg distribution

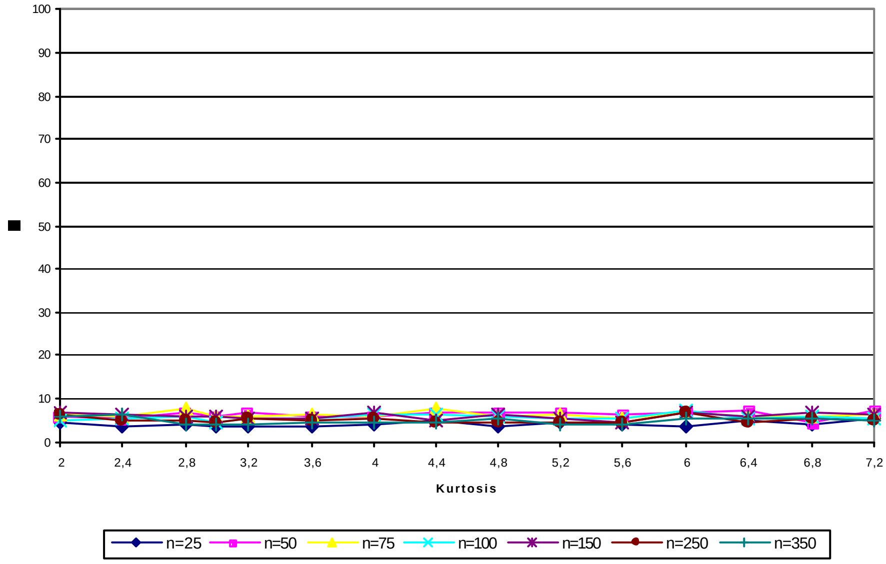

Figure 10: Power of X^2, PCS, PCK, PCR and JBK tests for equiprobable splits: Ramberg distribution (n=150)

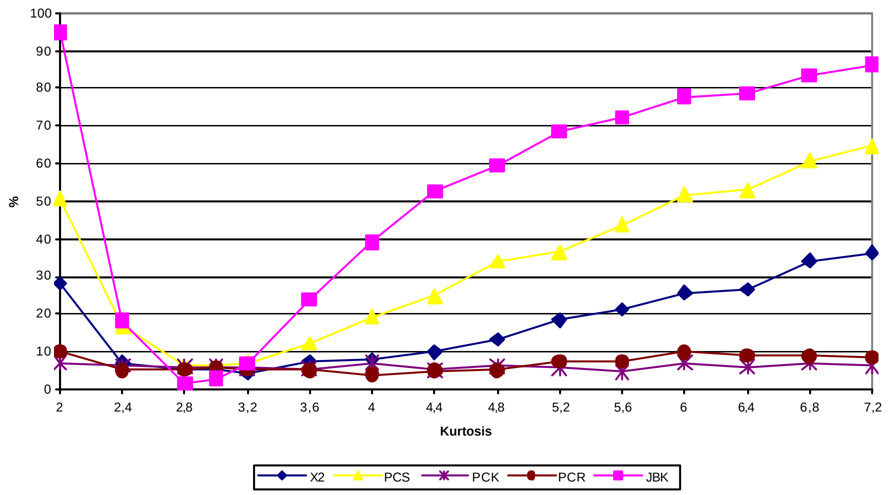

Figure 11: Intersection and equiprobable partition points for PCK test: Ramberg (kurtosis=6) and N(0,1)

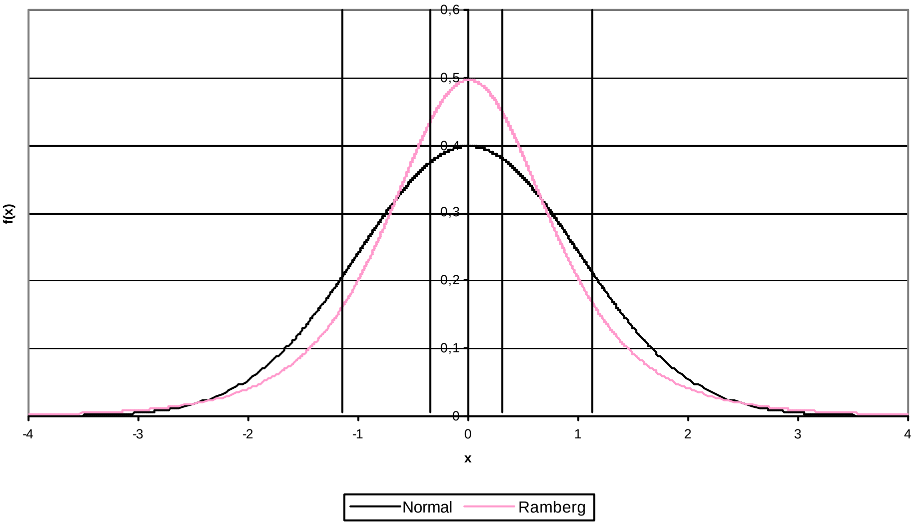

Figure 12: Power of the PCK test for non-equiprobable splits: Ramberg (kurtosis=6, n=150)

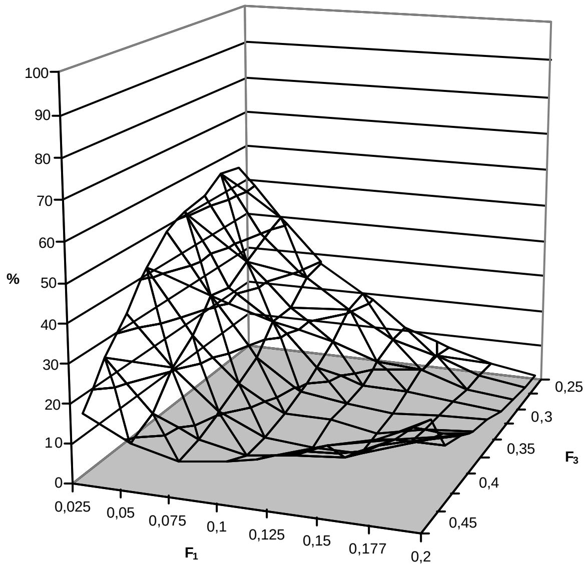

Figure 13: Power for the X^2, PCS and PCK tests for nonequiprobable splits: Ramberg distribution n=150

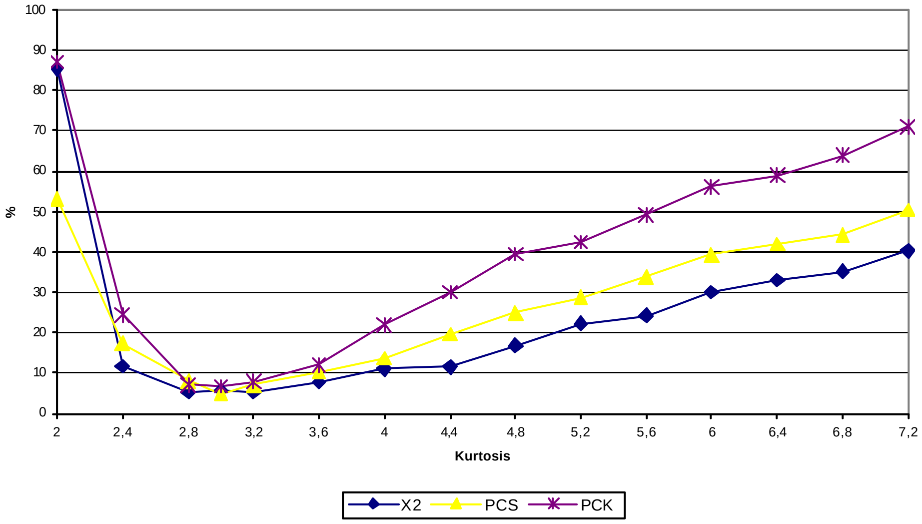

Figure 14: Power of the PCK test for equiprobable splits: Anderson's kurtotic distribution

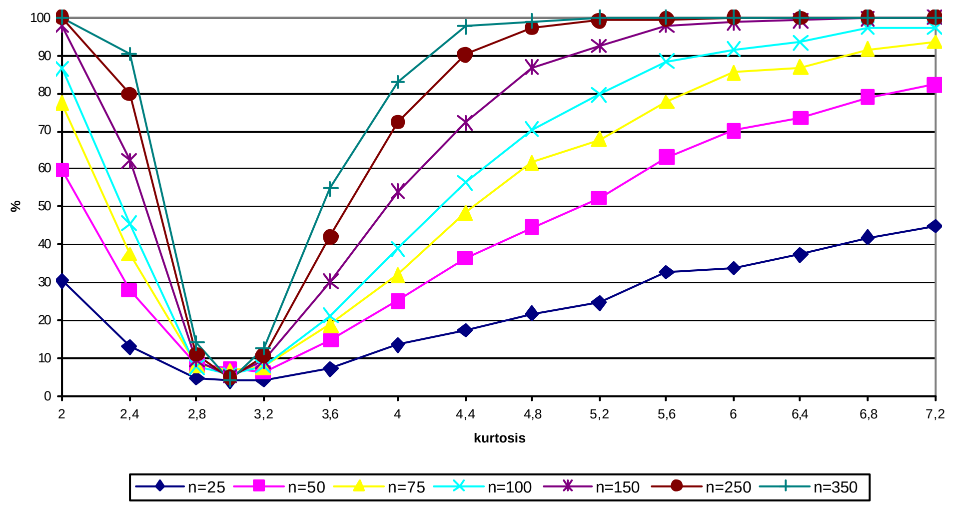

Figure 15: Power of the PCK test for non-equiprobable splits: Anderson (kurtosis=6, n=100)